AD-A068 147 TEXAS A AND M UNIV COLLEGE STATION INST OF STATISTICS F/G 12/1

QUANTILE FUNCTION VERSION OF BAYES THEOREM. (U)

MAR 79 E PARZEN DAAG29-78-G-0180

UNCLASSIFIED TR-A-4 ARO-16228.1-M NL

1 OF 1

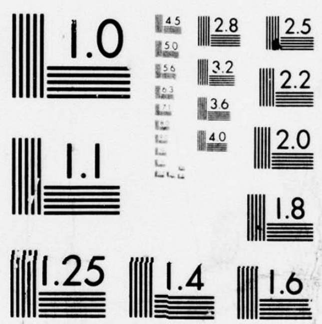
AD A068147

END

DATE FILMED

6 --79

DDC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

18 ARO 16228.1-M

BEST AVAILABLE COPY

12

TEXAS A&M UNIVERSITY
COLLEGE STATION, TEXAS 77843

Institute of Statistics
Phone 713 - 845-3141

QUANTILE FUNCTION VERSION OF BAYES THEOREM

by Emanuel Parzen
Institute of Statistics, Texas A&M University

Technical Report No. A-4
March 1979

Texas A & M Research Foundation
Project No. 3861

"Maximum Robust Likelihood Estimation and
Non-parametric Statistical Data Modeling"

Sponsored by the U.S. Army Research Office

Professor Emanuel Parzen, Principal Investigator

Approved for public release; distribution unlimited.

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER: Technical Report No. A-4	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle): (6) Quantile Function Version of Bayes Theorem.		5. TYPE OF REPORT & PERIOD COVERED: (9) Technical rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s): (10) Emanuel Parzen		8. CONTRACT OR GRANT NUMBER(s): (15) DAAG29-78-G-0180
9. PERFORMING ORGANIZATION NAME AND ADDRESS: Texas A&M University, Institute of Statistics, College Station, TX 77843		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS: Army Research Office, Research Triangle Park, NC 27709		12. REPORT DATE: (11) Mar 1979
		13. NUMBER OF PAGES: 26
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office): (12) 15p.		15. SECURITY CLASS. (of this report): Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA (14) TR-A-4

18. SUPPLEMENTARY NOTES

The findings in this report are not to be construed as an official Department of the Army position, unless so designated by other authorized documents.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Density-quantile function, dependence function, inference-distribution function, interquartile range, posterior quantile function, prior quantile function

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This paper proposes that prior and posterior distributions of a parameter be stated by its quantile function. It states a Bayes theorem for quantile functions: given data X and a likelihood function for X as a function of parameters $\theta_1, \ldots, \theta_k$, to each parameter θ_j one can associate a distribution function $D_j(u)$, $0 \le u \le 1$, such that $Q_{\theta_j|X}(u) = Q_{\theta_j}(D_j^{-1}(u))$

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

79 04 26 029

347 380

where Q_{θ_j} and $Q_{\theta_j|X}$ are respectively the prior and posterior quantile functions of θ_j. Plots of these functions provide graphical procedures for statistical data analysis.

QUANTILE FUNCTION VERSION OF BAYES THEOREM

by

Emanuel Parzen

Texas A&M University

Summary

This paper proposes that prior and posterior distributions of a parameter be stated by its quantile function. It states a Bayes theorem for quantile functions: given data X and a likelihood function for X as a function of parameters $\theta_1, \ldots, \theta_k$, to each parameter θ_j one can associate a distribution function $D_j(u)$, $0 \leq u \leq 1$, such that $Q_{\theta_j|X}(u) = Q_{\theta_j}(D_j^{-1}(u))$ where Q_{θ_j} and $Q_{\theta_j|X}$ are respectively the prior and posterior quantile functions of θ_j. Plots of these functions provide graphical procedures for statistical data analysis.

Some keywords: Density-quantile function, Dependence function, Inference-distribution function, Interquartile range, Posterior quantile function, Prior quantile function.

1. Introduction

Consider a continuous observation X (which could be a random vector) whose distribution depends on an unknown parameter θ to be estimated (in this section θ is a scalar). A model for X is specified by $f(X|\theta)$, the conditional probability density of X given θ. As a function of θ, for X fixed, one calls $f(X|\theta)$ the likelihood function.

When one adopts a Bayesian approach to the estimation of θ, one denotes it by $\underset{\sim}{\theta}$ when one wants to regard it as a random variable. The prior probability distribution of $\underset{\sim}{\theta}$ is described by a prior density, which we denote by $g(\theta)$, or by a prior distribution function denoted $G(\theta)$. Note that $G'(\theta) = g(\theta)$.

The prior distribution could represent either subjective probability or relative frequencies. The latter case occurs when one's prior distribution of $\underset{\sim}{\theta}$ is formed from a series of previously observed estimators $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n, \ldots$ of θ. This assumption provides a version of empirical Bayes estimation [see Bennett and Martz (1972)].

An important method of assessing a prior distribution is to assess the fractiles of the prior distribution [see Hays and Winkler (1971), p. 478]. A p fractile of the distribution of a continuous random variable X is a point c such that $F(c) = P(X \leq c) = p$. We call this concept the quantile function $Q(u)$, $0 \leq u \leq 1$, of X, and write

$$Q(u) = F^{-1}(u) = \inf\{x: F(x) \geq u\} .$$

The derivative $q(u) = Q'(u)$ is called the quantile-density function; it equals the reciprocal of the function

$$fQ(u) = f(Q(u)) = f(F^{-1}(u))$$

called the density-quantile function [see Parzen (1979)].

The prior quantile function of $\underset{\sim}{\theta}$ will be denoted $Q_{\underset{\sim}{\theta}}(u)$, $0 \leq u \leq 1$; it is defined by: $Q_{\underset{\sim}{\theta}}(u) = G^{-1}(u)$. The posterior quantile function $Q_{\underset{\sim}{\theta}|X}(u)$ of $\underset{\sim}{\theta}$ given X is defined as the inverse of the conditional

distribution function of $\underset{\sim}{\theta}$ given X, denoted

$$F_{\underset{\sim}{\theta}|X}(\theta) = P[\underset{\sim}{\theta} \leq \theta | X] , \qquad -\infty < \theta < \infty$$

Note that the posterior or conditional quantile function of $\underset{\sim}{\theta}$ given X includes as a special case the conditional median and quantiles of $\underset{\sim}{\theta}$ given X. Means and variances are also obtained directly from the quantile function:

$$\theta_0 = E[\underset{\sim}{\theta}] = \int_0^1 Q_{\underset{\sim}{\theta}}(u)\, du , \quad Var[\underset{\sim}{\theta}] = \int_0^1 \{Q_{\underset{\sim}{\theta}}(u) - \theta_0\}^2\, du$$

$$\theta^* = E[\underset{\sim}{\theta}|X] = \int_0^1 Q_{\underset{\sim}{\theta}|X}(u)\, du , \quad Var[\underset{\sim}{\theta}|X] = \int_0^1 \{Q_{\underset{\sim}{\theta}|X}(u) - \theta^*\}^2\, du$$

2. Inference-Distribution Functions and IQ

An approach to statistical inference in favor with all statisticians is to plot the likelihood function $f(X|\theta)$. This paper proposes that one plot in addition a distribution function $D(p)$, $0 \leq p \leq 1$, which can be regarded as an integrated likelihood (however, the integration takes place with respect to the argument u of the prior quantile function $Q_{\underset{\sim}{\theta}}(u)$). Define

$$d(u) = \frac{f(X|Q_{\theta}(u))}{\int_0^1 f(X|Q_{\theta}(u))\, du} .$$

which is a density function on $0 \leq u \leq 1$; define its distribution function

$$D(p) = \int_0^p d(u)\, du , \qquad 0 \leq p \leq 1$$

and quantile function $D^{-1}(u)$.

We call: d the inference-density of θ, D the inference-distribution function of θ, D^{-1} the inference-quantile function of θ (all given the data X).

The shape of D(p) will be an indicator of the relative importance of prior information and sample information. Two extreme cases are: (1) D(p) = p (a uniform distribution); (2) D(p) = 0 or 1 as $p < p_0$ or $p > p_0$ for some p_0 (a unit mass distribution). A quick and dirty measure of which of these situations prevails is provided by the interquartile range of D, denoted

$$IQ(\underset{\sim}{\theta}|X) = D^{-1}(0.75) - D^{-1}(0.25) \; ;$$

IQ close to .50 indicates a uniform distribution and IQ close to 0 indicates a unit mass distribution.

3. Bayes Theorem

Bayes theorem is usually regarded as: (1) a formula for the conditional probability density of $\underset{\sim}{\theta}$ given X:

$$f(\theta|X) = \frac{f(X|\theta)g(\theta)}{\int_{-\infty}^{\infty} f(X|\theta)g(\theta)d\theta} \tag{3.1}$$

and (2) a formula for the conditional mean of $\underset{\sim}{\theta}$ given X,

$$E[\underset{\sim}{\theta}|X] = \frac{\int_{-\infty}^{\infty} \theta f(X|\theta)g(\theta)d\theta}{\int_{-\infty}^{\infty} f(X|\theta)g(\theta)d\theta} \; . \tag{3.2}$$

The quantile version of Bayes theorem writes: (1) the formula for the conditional mean in the form

$$E[\underset{\sim}{\theta}|X] = \frac{\int_0^1 Q_\theta(u)\, f(X|Q_\theta(u))\, du}{\int_0^1 f(X|Q_\theta(u))\, du} \tag{3.3}$$

and, more importantly, (2) a formula for the conditional quantile function:

$$Q_{\underset{\sim}{\theta}|X}(u) = Q_{\underset{\sim}{\theta}}(D^{-1}(u)) \; , \qquad 0 \le u \le 1 \tag{3.4}$$

<u>Proof</u>: Write (3.2) as a Stieltjes integral:

$$E[\underset{\sim}{\theta}|X] = \frac{\int_0^1 \theta\, f(X|\theta)\, dG(\theta)}{\int_0^1 f(X|\theta)\, dG(\theta)} \; . \tag{3.5}$$

Make a change of variable $u = G(\theta)$, $\theta = Q_{\underset{\sim}{\theta}}(u)$ in (3.5); one obtains (3.3).

To prove (3.4), we first prove a relation between D(p) and the conditional distribution function $F_{\underset{\sim}{\theta}|X}(\theta)$:

$$F_{\underset{\sim}{\theta}|X}(Q_{\underset{\sim}{\theta}}(p)) = D(p) \; , \qquad 0 \le p \le 1 \; . \tag{3.6}$$

To prove (3.6), let $\theta_0 = Q_{\underset{\sim}{\theta}}(p)$ so that $G(\theta_0) = p$. Then, again making the change of variable $u = G(\theta)$, we can write

$$P[\underset{\sim}{\theta} \le \theta_0|X] = \int_{-\infty}^{\theta_0} f(\theta|X)\, d\theta$$

$$= \frac{\int_{-\infty}^{\theta_0} f(X|\theta)\, dG(\theta)}{\int_{-\infty}^{\infty} f(X|\theta)\, dG(\theta)} = \frac{\int_0^p f(X|Q_{\underset{\sim}{\theta}}(u))\, du}{\int_0^1 f(X|Q_{\underset{\sim}{\theta}}(u))\, du} \; ,$$

which is precisely equation (3.6).

To obtain (3.4) from (3.6), note that (3.6) implies

$$Q_{\underset{\sim}{\theta}|X}(D(p)) = Q_{\underset{\sim}{\theta}}(p) \; , \qquad 0 \le p \le 1 \tag{3.7}$$

A useful way to write (3.4) is:

$$Q_{\underset{\sim}{\theta}|X}(u) = Q_{\underset{\sim}{\theta}}(p) \tag{3.8}$$

where p satisfies

$$u = D(p) \quad \text{or} \quad p = D^{-1}(u) \; . \tag{3.9}$$

Bayes formula for quantile functions is easily implemented graphically.

(1) Draw graphs of $z = Q_{\underset{\sim}{\theta}}(p)$, and (given the data X) of $u = D(p)$

(2) Fix a value of u (for the median, $u = 0.5$); determine the value of p such that $u = D(p)$. Then $Q_{\underset{\sim}{\theta}|X}(u) = Q_{\underset{\sim}{\theta}}(p)$. In practice, one would consider computing $Q_{\underset{\sim}{\theta}|X}(u)$ for $u = 0.05, 0.10, 0.25, 0.50, 0.75, 0.90$, and 0.95. A suitable average of these values would approximate the conditional mean $E[\underset{\sim}{\theta}|X]$.

4. Example: Estimation of the mean of normally distributed data

To contrast the use of the Quantile function version of Bayes theorem with traditional versions, let us consider the estimation of the mean of normal random variables. Let $X_1, X_2, \ldots, X_n$ be independent normal with unknown mean θ and known variance σ^2. Let X denote the vector of observations $(X_1, \ldots, X_n)$. Then

$$f(X|\theta) = (2\pi\sigma)^{-n} \exp - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (X_i - \theta)^2$$

$$= (2\pi\sigma)^{-n} \exp - \frac{1}{2\sigma^2} \{ \sum_{u=1}^{n} (X_i - \bar{X})^2 + n(\bar{X} - \theta)^2 \} \tag{4.1}$$

where $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$.

The assumption often considered about the prior distribution of $\underset{\sim}{\theta}$ is that it is normal with mean θ_0 and variance σ_0^2. The prior density function is then given by

$$g(\theta) = \frac{1}{\sigma_0} \phi(\frac{\theta-\theta_0}{\sigma_0}) \; , \tag{4.2}$$

The Quantile approach would specify the prior quantile function of $\underset{\sim}{\theta}$ to be

$$Q_{\underset{\sim}{\theta}}(u) = \theta_0 + \sigma_0 \, \Phi^{-1}(u) \; , \tag{4.3}$$

Note that $\phi(x) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)$, $\Phi(x) = \int_{-\infty}^{x} \phi(y)dy$, and $\Phi^{-1}(u)$ is the quantile function of the standard normal distribution.

For normal data with a normal prior, one can explicitly calculate the posterior distribution of $\underset{\sim}{\theta}$ given X; it is normal,

$$f_{\underset{\sim}{\theta}|X}(y) = \frac{1}{\sigma[\underset{\sim}{\theta}|X]} \, \phi\left(\frac{y-E[\underset{\sim}{\theta}|X]}{\sigma[\underset{\sim}{\theta}|X]}\right) \; . \tag{4.4}$$

The conditional mean and variance of θ given X may be explicitly calculated and are given by

$$E[\underset{\sim}{\theta}|X] = \sigma^2[\theta|X](\theta_0\sigma_0^{-2} + \bar{X}(\sigma^2/n)^{-1})$$

$$\sigma^2[\underset{\sim}{\theta}|X] = (\sigma_0^{-2} + (\sigma^2/n)^{-1})^{-1} \tag{4.5}$$

Define the information numbers

$$I_0 = \sigma_0^{-2} \; , \quad I = (\sigma^2/n)^{-1} \; , \quad I[\underset{\sim}{\theta}|X] = \{\sigma^2[\underset{\sim}{\theta}|X]\}^{-1} \; . \tag{4.6}$$

Note that $I[\underset{\sim}{\theta}|X] = I_0 + I$. Next define parameters n_0 and IR by

$$\sigma_0^2 = \frac{\sigma^2}{n_0} \; , \quad IR = \frac{n_0}{n} = \frac{I_0}{I} \tag{4.7}$$

We call IR the information ratio; it represents the ratio of prior information to sample information, and is a useful parameter for expressing the conditional mean:

$$E[\underset{\sim}{\theta}|X] = \theta_0 + \frac{1}{1+IR}(\bar{X} - \theta_0) \ . \tag{4.8}$$

Some interpretations of this formula are as follows. If $IR \doteq 0$, then $E[\underset{\sim}{\theta}|X] \doteq \bar{X}$; if $IR \doteq \infty$, then $E[\underset{\sim}{\theta}|X] \doteq \theta_0$. A formula that is very useful for comparison purposes arises by expressing the estimator in terms of the deviation $h = \frac{\bar{X}-\theta_0}{\sigma_0}$:

$$\frac{E[\underset{\sim}{\theta}|X]-\theta_0}{\sigma_0} = \frac{1}{1+IR}\frac{\bar{X}-\theta_0}{\sigma_0} = \frac{1}{1+IR}h \tag{4.9}$$

The conventional version of Bayes Theorem has as its goal the calculation of $E[\underset{\sim}{\theta}|X]$ and credible intervals for $\underset{\sim}{\theta}$; as IR (the ratio of prior information to sample information) tends to 0 or ∞, $E[\underset{\sim}{\theta}|X]$ tends to the sample mean $\bar{X}$ or the prior mean θ_0 respectively. The quantile version of Bayes theorm has as its goal the calculation and plotting of $D(p)$ and $Q_{\underset{\sim}{\theta}|X}(u)$; limiting cases may be shown to be: as $IR \to \infty$, $D(p) = p$, the uniform distribution, and $Q_{\underset{\sim}{\theta}|X}(u) = Q_{\underset{\sim}{\theta}}(u)$; as $IR \to 0$, $Q_{\underset{\sim}{\theta}|X}(u)$ equals $\bar{X}$ for all u (that is, $F_{\underset{\sim}{\theta}|X}(\theta)$ has a unit probability mass at $\theta = \bar{X}$) and $D(p)$ is a purely discontinuous distribution function jumping from 0 to 1 (thus placing unit mass) at the value of p such that $Q_{\underset{\sim}{\theta}}(p) = \bar{X}$ (see Figure A).

The foregoing facts lead us to a conclusion with important practical applications: the shape of $D(p)$ is an indicator of the relative importance of prior information and sample information, and its interquartile range $IQ(\underset{\sim}{\theta}|X)$ can be interpreted as a generalization of IR (providing a measure of the ratio of prior information to sample information).

The quantile function version of Bayes theorem starts with computing and graphing $u = D(p)$. Since

$$f(X|\theta) \propto \exp\left(-\frac{n}{2\sigma^2}(\bar{X} - \theta)^2\right),$$

$$f(X|Q_{\theta}(u)) \propto \exp\left(-\frac{1}{2IR\sigma_0^2}(\bar{X} - \theta_0 - \sigma_0\,\phi^{-1}(u))^2\right) \ ,$$

d(u) can be expressed

$$d(u) = \frac{\exp\left(-\frac{1}{2IR}(h-\phi^{-1}(u))^2\right)}{\int_0^1 du\,\exp\left(-\frac{1}{2IR}(h-\phi^{-1}(u))^2\right)} \tag{4.10}$$

To compute $D(p)$, $0 \leq p \leq 1$, we can regard it as a function of h and IR. To permit other choices of prior quantile functions for $\underset{\sim}{\theta}$, we write the formula for $D(p)$ explicitly as follows:

$$D(p) = \frac{\int_0^p \exp\left(-\frac{1}{2IR}(h-Q_0(u))^2\right) du}{\int_0^1 \exp\left(-\frac{1}{2IR}(h-Q_0(u))^2\right) du} \tag{4.11}$$

where $Q_0(u) = \phi^{-1}(u)$.

The conditional median of $\underset{\sim}{\theta}$ given X, denoted $\theta^* = Q_{\theta|X}(0.5)$, is found by first finding the median p^* of $D(p)$; note that p^* is the value of p such that $D(p) = 0.5$. Then

$$\theta^* = Q_{\underset{\sim}{\theta}}(p^*) = \theta_0 + \sigma_0\,\phi^{-1}(p^*) \tag{4.12}$$

whence

$$\frac{\theta^*-\theta_0}{\sigma_0} = \phi^{-1}(p^*) \ . \tag{4.13}$$

This formula should be compared with the formula one obtains from the conventional version of Bayes theorem which is given by (9) but which we write in terms of θ^*:

$$\frac{\theta^*-\theta_0}{\sigma_0} = \frac{1}{1+IR}h \ . \tag{4.14}$$

The two formulas yield the same answer if and only if p^* satisfies

$$p^* = \Phi\left(\frac{1}{1+IR}\, h\right) \tag{4.15}$$

To compare numerically the conventional and Quantile versions of Bayes theorem, we computed for a range of values of IR and h, the function D(p), $0 \leq p \leq 1$, defined by (4.10). We verified that the value p^* at which $D(p) = 0.5$ is equal to $\Phi(h/(1 + IR))$. It shouldbe noted that we needed to consider only positive values of h in examining graphs of D(p) since, denoting it for clarity by $D_h(p)$, one may show that $D_{-h}(p) = 1 - D_h(1 - p)$.

To examine the effect of different choices of prior quantile functions for the parameter θ we considered also logistic and Cauchy prior quantile functions, which correspond to equation (4.11) with the choices

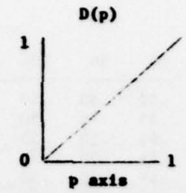
$$Q_0(u) = \log \frac{u}{1-u}\ , \quad \text{logistic;}$$

$$Q_0(u) = \tan \frac{\pi}{2}\,(2u - 1)\ , \quad \text{Cauchy;}$$

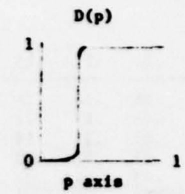
see Parzen [1979] for a listing of quantile functions.

Table A lists very approximate values of interquartile range $IQ(\theta|X)$ for the inference distribution function D(u) defined by (4.11) for different choices of h, IR, and Q_0; a study of the table will show IQ varies as a function of these parameters. Figure B graphs D(u) for $h = 1.5$, $Q_0 = \Phi^{-1}$, and various values of IR. These results illustrate the shapes of D and how IQ provides a measure of the relative weights of prior information and sample information in the statistical model under consideration.

Figure A

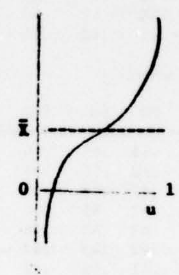


a uniform distribution for IR $\doteq \infty$



a unit mass distribution for IR $\doteq$ 0

$Q_{\theta|X}(u) = Q_\theta(u) = \theta_0 + \sigma_0\,\Phi^{-1}(u)$



$Q_{\theta|X}(u) = \bar{X}$

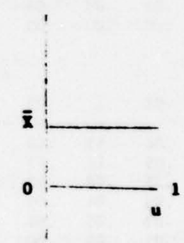


TABLE A

Approximate Values of $IQ(\underline{\theta}|X)$

Normal Prior

h | λ	.01	.1	.25	1	4	10	20	50	100	1000
0	.06	.16	.24	.37	.45	.48	.49	.50	.50	.50
.5	.06	.15	.22	.36	.45	.48	.49	.50	.50	.50
1.0	.05	.11	.18	.33	.45	.48	.49	.50	.50	.50
1.5	.03	.07	.12	.28	.44	.48	.49	.50	.50	.50
2.0	.03	.04	.07	.23	.42	.47	.49	.50	.50	.50
2.5	.03	.03	.03	.18	.40	.47	.49	.50	.50	.50
3.0	.03	.03	.03	.13	.39	.46	.49	.50	.50	.50

Logistic Prior

h | λ	.01	.1	.25	1	4	10	20	50	100	1000
0	.05	.11	.16	.28	.39	.44	.47	.48	.49	.50
.5	.04	.10	.15	.27	.39	.44	.47	.48	.49	.50
1	.03	.09	.14	.25	.38	.44	.47	.48	.49	.50
1.5	.03	.07	.11	.22	.37	.44	.47	.48	.49	.50
2	.03	.05	.08	.18	.35	.43	.46	.48	.49	.50
2.5	.03	.04	.06	.15	.33	.43	.46	.48	.49	.50
3	.03	.04	.05	.11	.31	.42	.46	.48	.49	.50
3.5	.03	.03	.04	.09	.28	.41	.46	.48	.49	.50
4	.03	.03	.03	.06	.25	.40	.45	.48	.49	.50

Cauchy Prior

h	.01	.1	.25	1	4	10	20	50	100	1000
0	.06	.13	.18	.27	.36	.40	.42	.45	.46	.49
.5	.05	.11	.17	.27	.35	.40	.42	.45	.46	.49
1	.05	.08	.13	.25	.35	.40	.42	.45	.46	.49
1.5	.03	.06	.09	.22	.35	.40	.42	.45	.46	.49
2.	.03	.05	.06	.17	.34	.40	.42	.45	.46	.49
2.5	.03	.03	.04	.12	.33	.40	.42	.45	.47	.49
3	.03	.03	.04	.09	.31	.39	.42	.45	.46	.49
3.5	.03	.03	.04	.07	.28	.39	.42	.45	.46	.49
4	.03	.03	.03	.05	.24	.38	.42	.45	.46	.49

5. Multiparameter Quantile version of Bayes theorem

Let the likelihood function of data X be $f(X|\theta_1, \ldots, \theta_k)$, a function of k parameters. The prior distribution of $\underline{\theta}_1, \ldots, \underline{\theta}_k$ is described in general by a joint density function $g(\theta_1, \ldots, \theta_k)$. The marginal distribution of a parameter θ_j is described by a density function $g_j(\theta_j)$, a distribution function $G_j(\theta_j)$, or a quantile function $Q_{\underline{\theta}_j}(u) = G_j^{-1}(u)$. The density-quantile function of $\underline{\theta}_j$ is $g_j Q_{\underline{\theta}_j}(u)$. The basic measure of dependence between $\underline{\theta}_1, \ldots, \underline{\theta}_k$ in the quantile domain is the <u>dependence</u> function

$$h(u_1, \ldots, u_k) = \frac{g(Q_{\underline{\theta}_1}(u_1), \ldots, Q_{\underline{\theta}_k}(u_k))}{g_1 Q_{\underline{\theta}_1}(u_1) \cdots g_k Q_{\underline{\theta}_k}(u_k)} ; \tag{5.1}$$

it is a density with domain the unit hypercube $0 \le u_j \le 1$, $j = 1, \ldots, k$ with the property that $\underline{\theta}_1, \ldots, \underline{\theta}_k$ are independently distributed if and only if $h(u_1, \ldots, u_k) \equiv 1$.

By a multiparameter quantile version of Bayes theorem we mean formulas for: (1) conditional distribution function $F_{\underline{\theta}_j|X}(y)$; (2) the conditional quantile function $Q_{\underline{\theta}_j|X}(u)$ of $\underline{\theta}_j$ given X; and (3) the conditional dependence function $h_{\underline{\theta}_1,\ldots,\underline{\theta}_k|X}(u_1,\ldots,u_k)$.

<u>Theorem</u>: There exist distribution functions $D_1(u), \ldots, D_k(u)$ such that

$$F_{\underline{\theta}_j|X}(Q_{\underline{\theta}_j}(u)) = D_j(u) ; \tag{5.2}$$

consequently

$$Q_{\underline{\theta}_j|X}(p) = Q_{\underline{\theta}_j}(u) \quad \text{where} \quad D_j(u) = p \tag{5.3}$$

or

$$\boxed{Q_{\underset{\sim}{\theta}_j|X}(p) = Q_{\underset{\sim}{\theta}_j}(D_j^{-1}(p))}. \tag{5.4}$$

An explicit formula for $D_j(u)$ is

$$D_j(u) = \int_0^1 du_1 \cdots \int_0^1 du_{j-1} \int_0^u du_j' \int_0^1 du_{j+1} \cdots \int_0^1 du_k \; c(u_1, \ldots, u_{j-1}, u_j', u_{j+1}, \ldots, u_k),$$

defining

$$b(u_1, \ldots, u_k) = f(X|Q_{\underset{\sim}{\theta}_1}(u_1), \ldots, Q_{\underset{\sim}{\theta}_k}(u_k))h(u_1, \ldots, u_k),$$

$$c(u_1, \ldots, u_k) = b(u_1, \ldots, u_k) + \int_0^1 \cdots \int_0^1 b(u_1', \ldots, u_k')du_1' \ldots du_k' .$$

<u>Proof</u>: These formulas follow immediately from a formula for the conditional joint distribution function of $\underset{\sim}{\theta}_1, \ldots, \underset{\sim}{\theta}_k$ given X, which we denote for brevity $F(\theta_1, \ldots, \theta_k|X)$:

$$F(\theta_1, \ldots, \theta_k|X) = \frac{\int_{-\infty}^{\theta_1} \cdots \int_{-\infty}^{\theta_k} f(X|\theta_1', \ldots, \theta_k')g(\theta_1', \ldots, \theta_k')d\theta_1' \ldots d\theta_k'}{\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(X|\theta_1', \ldots, \theta_k')g(\theta_1', \ldots, \theta_k')d\theta_1' \ldots d\theta_k'}$$

Change the variables of integration to u_j' defined by $\theta_j' = Q_{\underset{\sim}{\theta}_j}(u_j')$, and write $\theta_j = Q_{\underset{\sim}{\theta}_j}(u_j)$. Note that $d\theta_j' = \{g_j Q_{\underset{\sim}{\theta}_j}(u_j')\}^{-1} du_j'$. One obtains the <u>basic formula</u>

$$F(Q_{\underset{\sim}{\theta}_1}(u_1), \ldots, Q_{\underset{\sim}{\theta}_k}(u_k)|X) = \int_0^{u_1} \cdots \int_0^{u_k} c(u_1', \ldots, u_k')du_1' \ldots du_k'$$

To find $F_{\underset{\sim}{\theta}_j|X}(Q_{\underset{\sim}{\theta}_j}(u))$, let $u_1, \ldots, u_{j-1}, u_{j+1}, \ldots, u_k$ equal to 1, and $u_j = u$. The desired conclusions may now be inferred.

In order to take a posterior distribution of $\underset{\sim}{\theta}_1, \ldots, \underset{\sim}{\theta}_k$ as a new prior we need a formula for $h_{\underset{\sim}{\theta}_1, \ldots, \underset{\sim}{\theta}_k|X}(u_1, \ldots, u_k)$. One can show that

$$h_{\underset{\sim}{\theta}_1, \ldots, \underset{\sim}{\theta}_k|X}(u_1, \ldots, u_k) = \frac{c(D_1^{-1}(u_1), \ldots, D_k^{-1}(u_k))}{d_1(D_1^{-1}(u_1)) \ldots d_k(D_k^{-1}(u_k))} \tag{5.5}$$

where $d_k(u)$ is the derivative of $D_k(u)$.

The implications of the results of this section can only be realized by Bayesian statisticians who may desire to use the quantile function approach. It seems to be remarkable, however, that the relation between the posterior and prior quantile functions given by (5.4) holds even when there are many parameters. The influence of the joint likelihood function and the joint prior density is summarized in $D_j(p)$, $0 \le p \le 1$, a distribution function on the unit interval.

To estimate a parametric function $\phi = \phi(\theta_1, \ldots, \theta_k)$, one could use its conditional expectation $E[\phi|X]$; one can show that

$$E[\phi|X] = \int_0^1 \cdots \int_0^1 \phi(Q_{\underset{\sim}{\theta}_1|X}(u_1), \ldots, Q_{\underset{\sim}{\theta}_k|X}(u_k)) \; h_{\underset{\sim}{\theta}_1, \ldots, \underset{\sim}{\theta}_k|X}(u_1, \ldots, u_k)du_1 \ldots du_k \tag{5.6}$$

Note that the predictive density $f(X_{n+1}|X)$ of a new independent observation X_{n+1} given X is given by the expectation (with respect to the conditional distribution of $\underset{\sim}{\theta}_1, \ldots, \underset{\sim}{\theta}_k$ given X) of the likelihood function $f(X_{n+1}|\theta_1, \ldots, \theta_k)$.

This research was supported in part by the Army Research Office (Grant DAA29-78-G-0180).

References

Bennett, G. K. and Martz, H. F. (1972). A continuous empirical smoothing technique. Biometrika 59, 361-368.

Hays, W. L. and Winkler, R. L. (1971) Statistics: Probability, Inference, and Decision, Holt, Rinehart, and Winston.

Parzen, E. (1979). Nonparametric statistical data modeling. Journal American Statistical Assn. 74,

Figure B

Graphs of $D(p)$, $0 \leq p \leq 1$ for $h = 1.5$; IR = .01, .1, .25, 1, 4, 10, 20, 50, 100, 1000; $Q_0(u) = \Phi^{-1}(u)$ (normal prior), logistic prior, Cauchy prior.